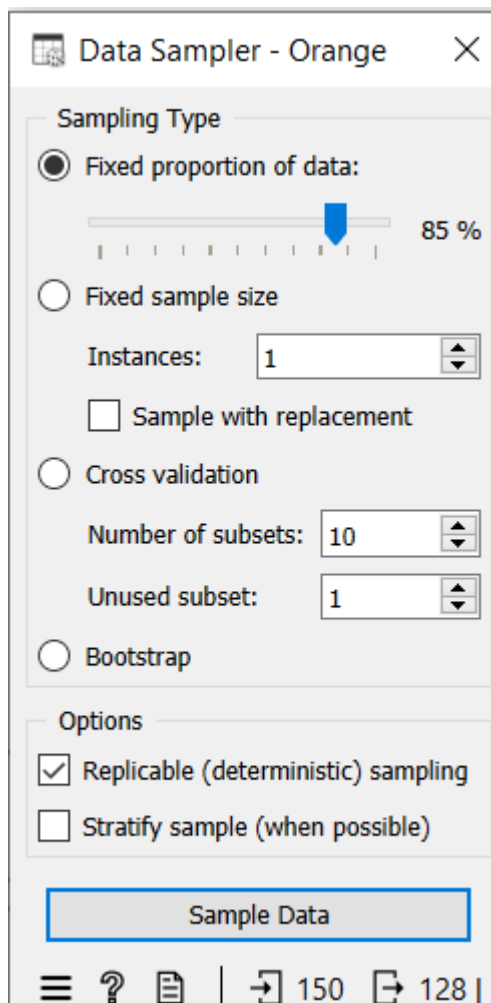


Pada sesi kali ini, fokus utamanya adalah membuat prediksi dengan logistic regression. Berikut adalah penjelasan dari komponen-komponen yang ada dalam gambar tersebut:

1. File: Menunjukkan langkah awal di mana dataset diimpor ke dalam sistem. Dari sini, data akan diproses lebih lanjut.
2. Data Table: Setelah data dimuat, pengguna dapat melihat dan mengeksplorasi data dalam bentuk tabel, yang memungkinkan analisis awal terhadap atribut-atribut yang ada.
3. Scatter Plot: Komponen ini digunakan untuk memvisualisasikan hubungan antara dua variabel dalam dataset, memungkinkan pengguna untuk mengidentifikasi pola atau korelasi yang mungkin ada.
4. Data Sampler: Ini adalah alat yang digunakan untuk membagi dataset menjadi dua bagian: data pelatihan dan data pengujian. Pembagian ini penting untuk mengembangkan model yang dapat diuji efektivitasnya.
5. Test and Score: Setelah model dibangun, langkah ini digunakan untuk menguji model terhadap data pengujian dan mengevaluasi kinerjanya berdasarkan metrik tertentu.
6. Logistic Regression: Menunjukkan penerapan algoritme regresi logistik untuk memodelkan data. Ini adalah metode yang umum digunakan untuk prediksi biner, seperti prediksi risiko stroke dalam konteks ini.
7. Predictions: Langkah terakhir dalam alur ini adalah menghasilkan prediksi berdasarkan model yang telah dibangun, memberikan hasil yang dapat digunakan untuk analisis lebih lanjut atau pengambilan keputusan.



1. Sampling Type:

Fixed proportion of data: Opsi ini memungkinkan pengguna untuk menentukan persentase data yang akan diambil dari dataset asli. Dalam gambar, persentase yang ditetapkan adalah 85%, yang berarti 85% dari data akan digunakan untuk tujuan sampling.

2. Fixed sample size:

Pengguna dapat mengatur ukuran sampel tetap dengan menentukan jumlah instance yang diinginkan. Ini berguna jika pengguna ingin mengambil sampel dengan jumlah tertentu, terlepas dari ukuran dataset asli.

3. Instances:

Ini adalah pengaturan untuk menentukan jumlah instance (baris data) yang akan diambil dari dataset.

4. Cross validation:

Opsi ini memungkinkan pengguna untuk melakukan validasi silang, yang merupakan teknik untuk mengevaluasi model dengan membagi data menjadi beberapa subset, melatih model pada sebagian data, dan mengujinya pada subset yang lain.

5. Number of subsets:

Jika validasi silang dipilih, pengguna dapat menentukan jumlah subset yang akan dibuat. Dalam gambar, jumlah subset yang ditetapkan adalah 10.

6. Unused subset:

Menunjukkan pengaturan untuk subset yang tidak digunakan dalam proses pelatihan, yang bisa digunakan untuk validasi model.

7. Bootstrap:

Opsi ini memungkinkan pengguna untuk melakukan sampling dengan penggantian, yang berarti data yang sama dapat dipilih lebih dari sekali dalam sampel.

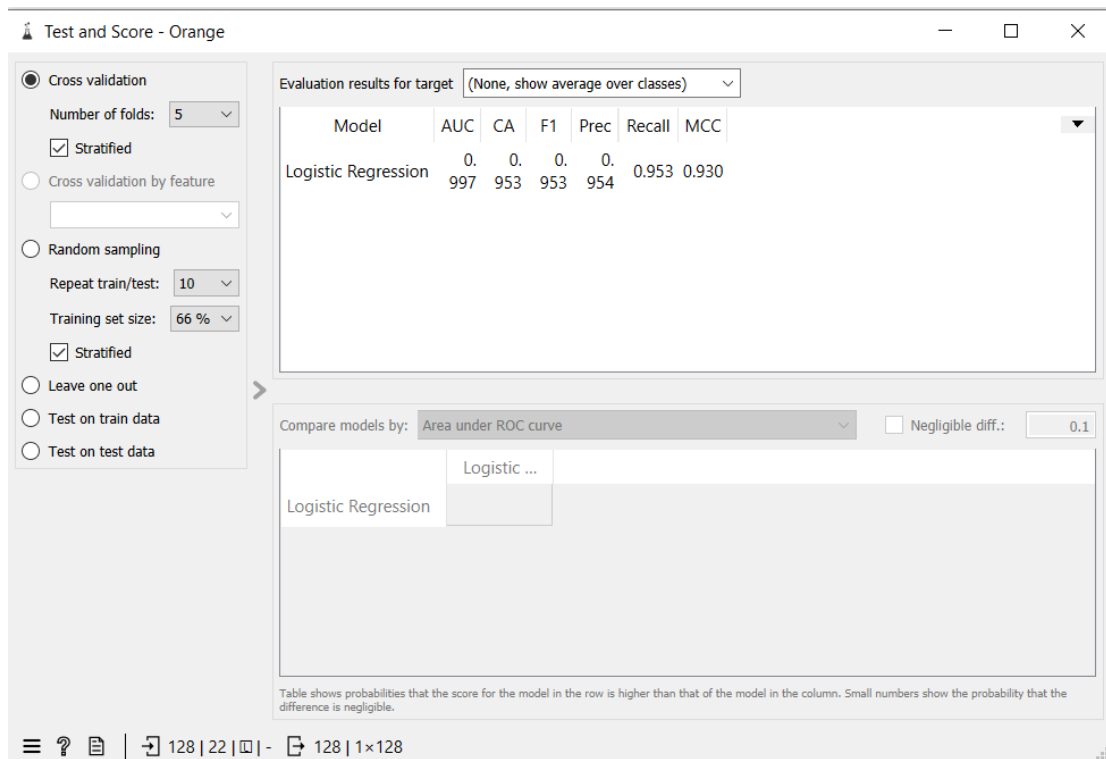
8. Options:

Replicable (deterministic) sampling: Dengan mengaktifkan opsi ini, pengguna memastikan bahwa hasil sampling dapat direproduksi di masa mendatang, menghasilkan sampel yang sama setiap kali dijalankan.

Stratify sample (when possible): Opsi ini memastikan bahwa sampel mencerminkan proporsi kelas yang ada dalam dataset asli, yang penting dalam analisis data yang tidak seimbang.

9. Sample Data:

Tombol ini digunakan untuk memulai proses pengambilan sampel berdasarkan pengaturan yang telah ditentukan.



1. Model:

Nama model yang dievaluasi, dalam hal ini adalah Logistic Regression.

2. AUC (Area Under the Curve):

Metrik yang menunjukkan performa model dalam klasifikasi biner. Nilai AUC mendekati 1 menunjukkan model yang sangat baik, sedangkan nilai mendekati 0.5 menunjukkan performa yang buruk. Di sini, AUC adalah 0.997, yang menunjukkan model berkinerja sangat baik.

3. CA (Classification Accuracy):

Menunjukkan persentase prediksi yang benar dari total prediksi. Di sini, CA adalah 0.953, yang berarti model benar dalam 95.3% kasus.

4. F1:

Metrik yang menggabungkan presisi dan recall, memberikan gambaran tentang keseimbangan antara keduanya. Di gambar, nilai F1 adalah 0.953.

5. Prec (Precision):

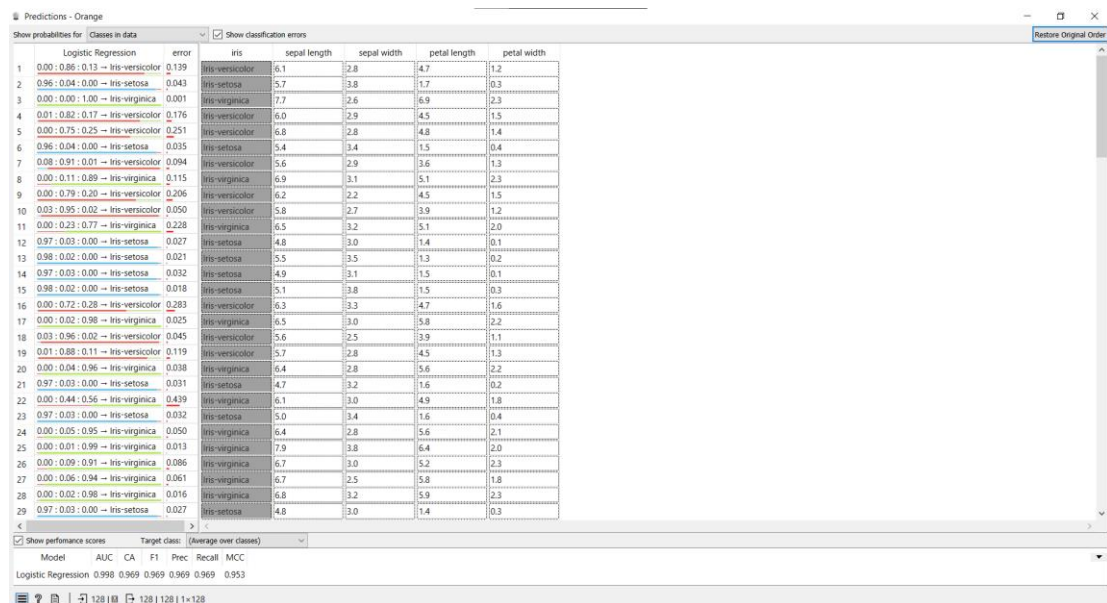
Menunjukkan proporsi prediksi positif yang benar dari semua prediksi positif. Di sini, precision adalah 0.953.

6. Recall (R):

Menunjukkan proporsi prediksi positif yang benar dari semua kasus positif yang sebenarnya. Nilai recall juga adalah 0.953.

7. MCC (Matthews Correlation Coefficient):

Metrik yang memberikan gambaran menyeluruh tentang kualitas klasifikasi biner, dengan nilai berkisar antara -1 dan 1. Di sini, MCC adalah 0.930.



The screenshot shows the 'Predictions - Orange' window. It displays a table with columns for 'Logistic Regression', 'error', 'Iris', 'sepal length', 'sepal width', 'petal length', and 'petal width'. The table contains 29 rows of data. Below the table, there is a section for 'Show performance scores' with a 'Target class: (Average over classes)' dropdown. The performance metrics are listed as follows:

Model	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	0.998	0.969	0.969	0.969	0.969	0.953

1. Model Predictions:

Output ini mencakup kolom yang menunjukkan prediksi yang dihasilkan oleh model untuk setiap instance dalam dataset. Misalnya, kolom "Predictions" menunjukkan kelas yang diprediksi oleh model, sedangkan kolom "Actual" menunjukkan kelas yang sebenarnya.

2. Error Analysis:

Kolom "Error" mengindikasikan perbedaan antara prediksi dan nilai aktual. Ini membantu mengidentifikasi kesalahan yang dibuat oleh model, yang penting untuk evaluasi kinerja.

3. Features:

Fitur-fitur yang digunakan untuk membuat prediksi dicantumkan, seperti:

- Sepal Length: Panjang sepal.
- Sepal Width: Lebar sepal.

- Petal Length: Panjang petal.
- Petal Width: Lebar petal.

Setiap fitur memiliki nilai yang terukur untuk setiap instance dalam dataset, yang memberikan konteks untuk prediksi yang dibuat.

4. Performance Metrics:

Di bagian bawah antarmuka, metrik kinerja model ditampilkan, memberikan gambaran umum tentang seberapa baik model melakukan prediksi.

- AUC: Nilai AUC yang tinggi menunjukkan kemampuan model untuk membedakan antara kelas dengan baik.
- Accuracy: Persentase prediksi yang benar, yang menunjukkan seberapa sering model memberikan prediksi yang akurat.
- F1 Score: Menggambarkan keseimbangan antara presisi dan recall, memberikan informasi yang lebih komprehensif tentang kinerja model.
- Precision dan Recall: Metrik ini membantu memahami seberapa baik model dalam memprediksi kelas positif dan seberapa banyak dari kelas positif yang berhasil diidentifikasi.